International Association of Societies of Design Research Congress 2023
LIFE-CHANGING DESIGN
Politecnico di Milano
9-13 October 2023

Designing interfaces for text-to-image prompt engineering using stable diffusion models: a human-AI interaction approach

Kim, Seonuk[a]; Ko, Taeyoung[a]; Kwon, Yousang[a]; Lee, Kyungho*[a]

[a] Expressive Computing Lab., Department of Design, UNIST, Ulsan, Republic of Korea
* kyungho@unist.ac.kr

doi.org/10.21606/iasdr.2023.448

The use of generative artificial intelligence (AI) is more vital ever than before for creating new content, especially images. Recent breakthroughs in text-to-image diffusion models have shown the potential to drastically change the way we approach image content creation. However, artists still face challenges when attempting to create images that reflect their specific themes and formats, as the current generative systems, such as Stable Diffusion models, require the right prompts to achieve the desired artistic outputs. In this paper, we propose future design considerations to develop more intuitive and effective interfaces that can be used for text-to-image prompt engineering from a human-AI interaction perspective using a data-driven approach. We collected 78,911 posts from the internet community and analyzed them through thematic analysis. Our proposed directions for interface design can help improve the user experience as well as usability, ultimately leading to a more effective and desired image generation process for creators.

Keywords: stable diffusion; human-ai interaction; thematic analysis; interface design

1 Introduction

The use of generative artificial intelligence (AI) is more vital than ever before for creating new content, including images, codes, music, text, and videos. Recent breakthroughs in text-to-image diffusion models have shown the potential to drastically change the way we approach image content creation drastically. The results seem to exceed human capability using various diffusion models. After Stable Diffusion was released as open source, various derivative models appeared, as Stable Diffusion consumes fewer computing resources and is easy to fine-tune (Rombach et al., 2022). NovelAI Diffusion, one of them, generated high-quality anime images and caught the attention of creators and artists.

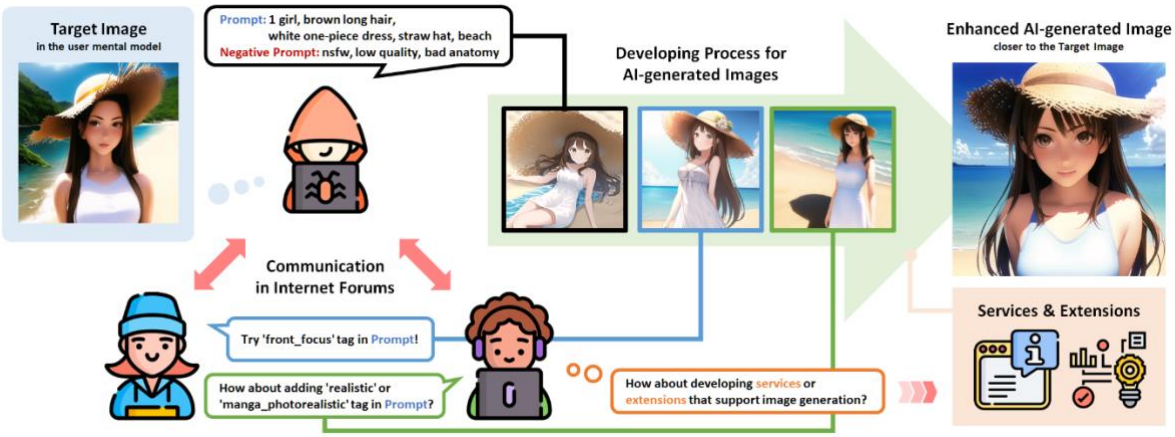

Figure 1. This figure illustrates how users in an internet forum collaborate to learn and develop their own approach in prompt engineering with Stable Diffusion models. They strive to create the exact image they wish to see through trials, errors, and communication.

The diffusion model generates images from text based on the principle of an autoencoder. An autoencoder consists of an encoder and a decoder, where the encoder vectors the user-input text into an AI-understandable format, and the decoder outputs the embedding vector in a human-understandable format, such as an image. Trained on data with matched texts and images, Stable Diffusion uses the embedding vector created by the encoder as a cue to vectorize, denoise, and recover the image so that it outputs an image that intentionally reflects the input from the user, regardless of the information in the original image. In other words, the image the AI thinks it has recovered is actually a completely new image created to reflect the information the user intended.

Since Stable Diffusion was open source, it became a game-changer for AI image generation models. Although there have been text-to-image models that generate high-quality images before, such as DALL·E 2 and Midjourney, they are not open source, and the variables that can be manipulated are limited to the input text, limiting their application. Stable Diffusion, on the other hand, can not only analyze user input text but also fine-tune models and LoRAs to drive the desired image style, such as anime or realistic and pose of the character, and even improve the graphical user interface (GUI) with tools like Gradio.

However, much of this is only customizable by developers, and in the long run, text prompts are arguably the most important factor for creators using various image generative AI services. Creators without technical backgrounds still face challenges when attempting to create images that reflect their specific intentions and styles, as the current generative systems, such as Stable Diffusion models, require the right prompts to achieve the desired artistic outputs. In addition, there are only a few works on what inconveniences users face in finding text prompts to use as input for image generative AI using the diffusion model, what user interface components are essential, and how the design should be improved for the service. This is because the current image generation process is based on the users' text inputs, engineered prompts. The purpose of prompt engineering is to design prompts that elicit the most relevant and desired response from a Large Language Model (LLM). It needs a deep understanding of the capabilities of the model and crafting prompts that will effectively utilize them.

To address the gap, we propose future design considerations to develop more intuitive and effective interfaces that can be used for text-to-image prompt engineering from a human-AI interaction

perspective using a data-driven approach. We considered that in order to propose an improvement direction for the interface design that applies generative AI as new design materials, it is necessary to identify the relationship between the user behavior patterns and services. Therefore, we selected the following research questions:

1. What behavior patterns do users of image generative AI services exhibit to find prompts?
2. What is the relationship between each behavior pattern and the sequence of processes?
3. What improvements to the extension and service are needed through user-developed support services to facilitate specific behavior patterns or reduce unnecessary procedures?

We chose the Arcalive AI Art Channel, a Korean internet forum where anime image generation using AI is actively discussed (like r/StableDiffusion on Reddit), as the site to explore. We collected 78,911 posts and analyzed 1,068 posts with more than 10 recommendations (voted up) that were recognized as valuable by other users. The posts we focused on were highly representative of the prompt engineering and trial-and-error process experienced by users, and closely related to interface design.

Through the analysis, we identified and classified the patterns with 25 behaviors and 8 types of users that show the collaborative process to make the best use of stable diffusion models to create the image outputs that meet the creators' requirements. Our study provides valuable insights into the challenges faced by users when generating images with text-to-image models. Our proposed directions for interface design can help reduce unnecessary procedures and improve the user experience, ultimately leading to a more effective and desired image generation process for creators without strong AI backgrounds.

2 Related work

2.1 Design for human-AI interaction

In the context of products and services based on human-AI interaction, AI has been discussed as *"a new and difficult design material"* in UX design because it requires a completely new paradigm shift in its use and necessitates background knowledge and experience in it workings (Yang et al., 2018, 2020). Software engineers have used Gradio (Abid et al., 2019) to create new interface design-based demo services that fit the diffusion model, such as the current Stable Diffusion web UI (AUTOMATIC1111, 2022/2023), and launched them as completed services, such as DALL·E 2, Midjourney, and NovelAI Image Generation. In this paper, we propose introducing a recommendation system of extension programs and digital drawing interfaces with a high degree of freedom to maximize the user experience regarding image generation control.

2.2 Interpret and control the diffusion model

DAAM (Tang et al., 2022) visualized how each tag the users input into the prompt affects which part of the image is generated. ControlNet (Zhang & Agrawala, 2023) presented a way for users to input not only text but also boundaries, poses, segments, depth, and others to control the generated image. In this paper, we propose directions for interface design that improve the drawing interface to provide control and user experience, such as retouching to generate desired images, and to facilitate collecting training data for fine-tuning via user interaction.

3 Method and results

3.1 Method and subjects

The Arcalive AI Art channel is a Korean internet forum dedicated to the topic of AI-generated images with over 36,000 subscribers. The first post was made on September 8, 2022, and as of February 15, 2023, there were more than 120,000 posts on the channel. While the forum initially covered a variety of AI-generated images, most of the posts now focus on anime images. We selected 1,068 posts with at least 10 recommendations as of November 14, 2022, and conducted the study in three steps. First, we read all the selected posts and conducted a case study on users of the AI anime image generation services. Second, we identified the searching process of users for prompts to generate the desired images and defined behaviors and types of users through thematic analysis based on the case study (Braun & Clarke, 2006). Third, we analyzed which types were supported by the services that users voluntarily developed or proposed for convenience, and further analyzed the pros and cons of each service to suggest ways to maximize the user experience of creating the desired images based on the identified types. The types defined by thematic analysis are important because they increase confidence in providing quantitative results and allow correlation to be checked. Furthermore, it allows case studies focused on user behavior, extensions, and services to verify that the correlations identified are causal by screening posts in forums where images are important, as well as text.

3.2 Case studies

In our case study, we found that most users in the forum were using generative AI for fun, such as showing off the images they generated and sharing prompts with other users. However, there are a few users whose behavior deviates from these trends. For instance, some users have been shown to do prompt engineering with DAAM (Tang et al., 2022) via an extension to the Stable Diffusion web UI to confirm which components of the generated image each tag affects or control experiments with prompt matrices and X/Y/Z plots. On the other hand, some users have developed and fine-tuned their own models and hypernetworks to fit their intentions, sharing and organizing them in forums. Recently, Stable Diffusion web UI has been updated to provide more convenient features and extensions, and the training methods shared with other users via posts on the forums have lowered the barrier to entry for non-AI experts to take the lead in fine-tuning on the forums.

3.3 Thematic analysis

We summarized each post around its content and documented it with URLs. We focused on the actions users take to explore the prompts and attempted to categorize them based on the documentation. During the categorization process, the content underlying each behavior was extracted and aggregated, and 25 behaviors were defined. Based on the 25 behaviors, we defined 8 types of users as shown in Table 1. We focused on how each behavior can be generalized. In the explorer type, there is a significant difference between the behavior of exploring the meaning of tags and the behavior of viewing tags as symbols and intentionally strengthening or weakening the meaning using weights, so we classified them once more according to the approach. Moreover, we classified the types of the 1,321 posts that received 10 or more recommendations as of November 30, 2022. As shown in Table 1, the proportion of consumer and producer types is more than 60%, while the rest of the types, except for the semantic approach explorer type, are not active below about 20%. Since posts with a high number of recommendations will show results in Table 1, and informational posts will receive a high number of recommendations, we can infer that generating and sharing AI-

generated images are a larger part of AI art forum activity than the researchers found. Furthermore, we found that the types of each behavior pattern have the relationship shown in Figure 3 based on the qualitative analysis in which researchers cross-validated user behavior procedures, including user behaviors observed in each post.

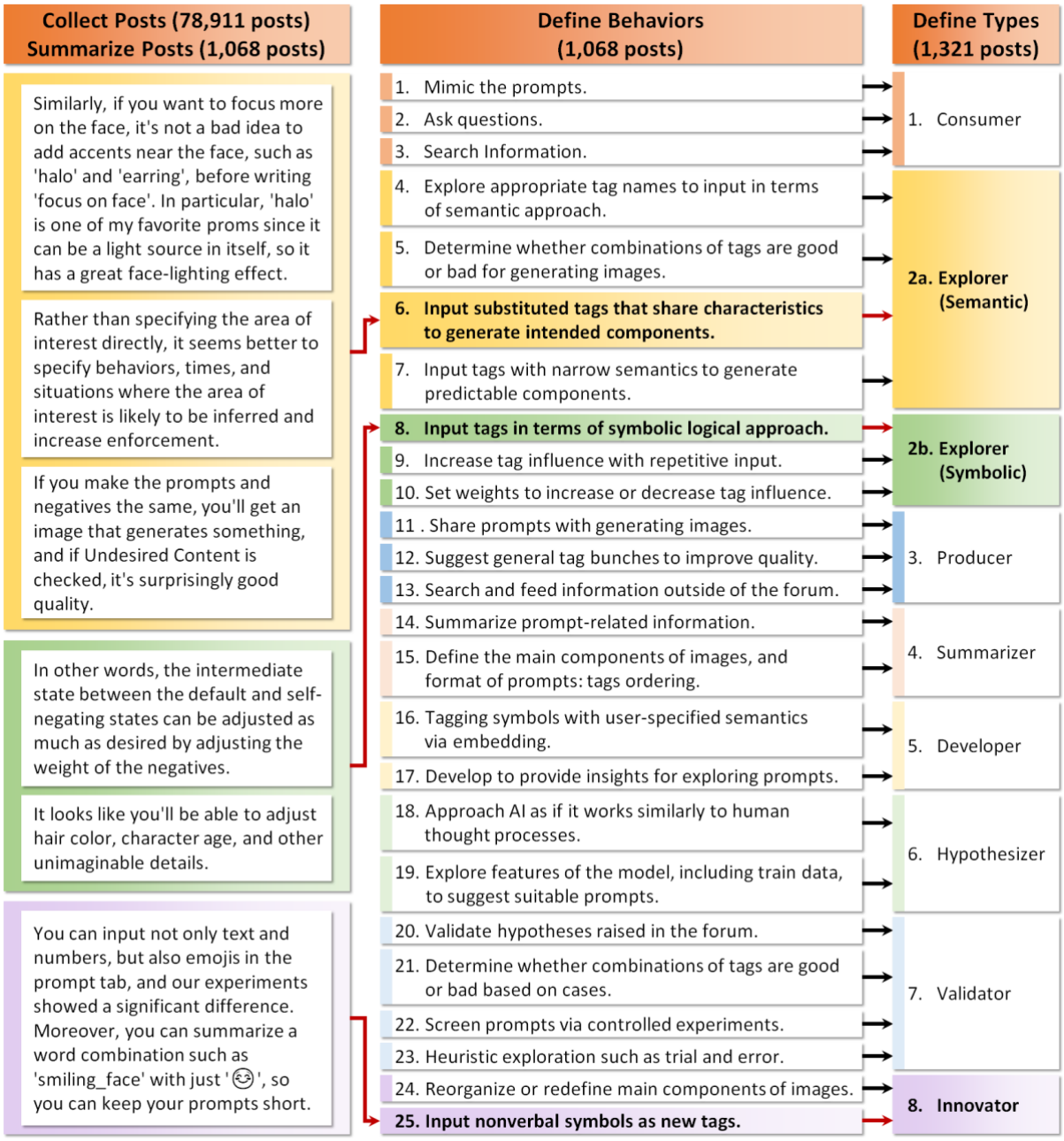

Figure 2. The diagram shows the process of thematic analysis to define 8 types. First, 1,068 posts that received 10 or more recommendations on the Arcalive AI Art channel, the internet forum, out of collected 78,911 posts until November 14, 2022, were screened, read, excerpted, and summarized, and 25 behaviors were defined through discussions among the researchers. In this figure, the texts included in the Screen, Read, Excerpt, and Summarize Posts are excerpted versions of the actual posts translated from Korean to English. After defining 25 behaviors, we further defined 8 types through discussions among the researchers. For Explorers, we further categorized them into semantic and symbolic approaches based on their approach. Then, 1,321 posts that received 10 or more recommendations on the Arcalive AI Art channel until November 30, 2022, were collected in the same way as November 14, 2022, and quantitatively analyzed by matching the types defined with user behavior in Table 1.

Table 1. The table shows the results of the thematic analysis of user behavior exploring the prompts

Types	**Approach**		**Behaviors**	***N***	***%***
Consumer		1	Mimic the prompts.		
		2	Ask questions.	856	64.8
		3	Search for information.		
Explorer	Semantic	4	Explore appropriate tag names to input in terms of the semantic approach.		
		5	Determine whether combinations of tags are good or bad for generating images. (*e.g., some components were not generated by stronger semantic tags.*)	473	35.8
		6	Input substituted tags that share characteristics to generate intended components.		
		7	Input tags with narrow semantics to generate predictable components.		
	Symbolic	8	Input tags in terms of symbolic logical approach. (*e.g., control the hair color by inputting weighted tags at the negative prompt.*)	107	8.1
		9	Increase tag influence with repetitive input.		
		10	Set weights to increase or decrease tag influence.		
Producer		11	Share prompts with generating images.		
		12	Suggest general tag bunches to improve quality.	892	67.5
		13	Search and feed information outside of the forum.		
Summarizer		14	Summarize prompt-related information.	160	12.1
		15	Define the main components of images and format of prompts: tag ordering.		
Developer		16	Tagging symbols with user-specified semantics via textual inversion; also called embedding in the forum.		
		17	Develop to provide insights for exploring prompts. (*e.g., train new models, tune hypernetworks, and develop support services.*)	266	20.1
Hypothesizer		18	Approach AI as works similarly to human thought processes.		
		19	Explore features of the model, including train data, to suggest suitable prompts.	22	1.7
Validator		20	Validate hypotheses raised in the forum.		
		21	Determine whether combinations of tags are good or bad based on cases.	137	10.4
		22	Screen prompts via controlled experiments.		
		23	Heuristic exploration such as trial and error.		
Innovator		24	Reorganize or redefine the main components of images. (*e.g., input alpha transparency as a tag into the prompt.*)	60	4.5
		25	Input nonverbal symbols as new tags. (e.g., special symbols such as Unicode, emoji, etc.)		

In this table, N is the number of posts, and % is the percentage of posts in each type out of the total 1,321 posts that received 10 or more recommendations on the Arcalive AI Art channel, the internet forum, until November 30, 2022. The analysis categorized user behavior into 8 types and utilized both quantitative and qualitative methods. These findings can be applied to improve the interface design for AI-generated art.

3.4 Extensions of Stable diffusion web UI

The Stable diffusion web UI provides extensions to help users generate images in conjunction with external sites. These extensions act similarly to plugins for software tools. Researchers selected seven extensions that have received a lot of recommendations in the forums, showing which type they fit into and their descriptions in Tables 2 and 3, respectively. The uppercase letters in the Services row are the first letter of each type name, indicating the corresponding type. Fully filled circles mean that the service has the nature of that type. However, the explorer type is divided into semantic and symbolic approaches, so these were labeled by Se and Sy, respectively.

Among these extensions, ControlNet and DAAM had a particularly strong impact on users. With the advent of ControlNet, the results of users creating short videos with each image as a frame became noticeably more stable. With the advent of DAAM, users could clearly verify tags, discovering that tags like 8k wallpaper and unity, which they had been convinced were effective, were dummy tags.

Table 2. The table presents the names of the extensions that are provided by the Stable Diffusion web UI, along with the type that corresponds to each extension. Most of these extensions have been developed by users to support specific tasks in the generation process.

Extensions	**C**	**Se**	**Sy**	**P**	**S**	**D**	**H**	**V**	**I**
ControlNet (Zhang & Agrawala, 2023)	●	●		●		●	●	●	●
DAAM (Tang et al., 2022)	●	●		●		●	●	●	●
InstructPix2Pix (Brooks et al., 2023)	●	●	●	●			●	●	●
Ddetailer (dustysys, 2022/2023)	●	●					●	●	
Latent Couple Extension (opparco, 2023/2023)	●	●	●			●	●	●	●
Dynamic Prompts (Eyal, 2022/2023)	●	●	●	●		●	●	●	
Supermerger (hako-mikan, 2023/2023)	●	●				●	●	●	

Table 3. The table provides descriptions of Stable Diffusion web UI extensions that support the generation of AI images, informed by recommended posts in the Arcalive AI Art channel, an internet forum.

Extensions	**Descriptions**
ControlNet	Add OpenPose, Depth, Segment, Canny, and more as inputs to help generate images with more precise pose, motion, and composition.
DAAM	Visualize the color of areas in the image that were contributed to generate each of the tags used.
InstructPix2Pix	To overcome the disadvantage of not knowing what text prompt the generated figure was generated from, so that the figure can be modified in a similar form, the generated figure can be modified with additional narrative input. (*e.g., To robotically turn the generated image of the statue of David, just add the additional input 'turn him into a cyberg'. Before, we would have had to regenerate the image by adding another tag, keeping the text and parameter settings that generated it.*)

ddetailer	Upscale the resolution of the face of specific people or characters while keeping the background intact to generate a more accurate and intended image.
Latent Couple Extension	Support for generating characters with two or more different features in a single image with ease and good quality. (e.g., users can specify different accessories for Character A and B although generate one image at once.)
Dynamic Prompts	When the user has a bunch of tags in sets A, B, C, and so on, it helps generate images repeatedly using a randomized search that randomly selects one of the tags that corresponds to each set the user has enabled.
Supermerger	Synthesis-based fine-tuning, such as model checkpoint synthesis, allows users to visualize and compare the process of creating new image styles.

3.5 Creativity support services

Creativity support services that users voluntarily developed or proposed for convenience were identified during the case study, and the corresponding types for each service are shown in Table 4. Descriptions of these services can also be seen in Table 5. The uppercase letters in the Services row are the first letter of each type name, indicating the corresponding type. The structure of the tables is the same for Tables 2 and 4, and Tables 3 and 5, respectively.

In particular, we confirmed the promotion of the trend from consumer to producer since the AI Image EXIF Viewer allowed users to see the prompt information or be recommended similar prompt tags through the integrated DeepDanbooru when no information was attached just by clicking on the image inside the forum. We also confirmed demands for the service that matches images and tags and downloads normalized images for train data or cloud GPU rental for image generation, and image retouching through user comments on the post. For example, we often found users with limited computing resources or utilization using Discord Image Bot for image generation, and we also found that posts providing information on using cloud GPU rental services like RunPod to fine-tune models for users in the developer type received a lot of recommendations.

Table 4. The table presents the names of the services that users have developed to support specific tasks in the generation process, along with the type that corresponds to each service.

Services	**C**	**Se**	**Sy**	**P**	**S**	**D**	**H**	**V**	**I**
Snowshell (YukihoAA, 2017/2023)				●					
Illuminarty	●					●	●	●	
AI Image EXIF Viewer	●	●	●	●				●	
NovelAI Tag Generator	●	●		●	●				
WebUI Tag Autocomplete	●	●		●					
AIBooru	●	●	●	●	●		●	●	
Prompt Search	●	●	●	●	●		●	●	
NovelAI Tag Experiments	●	●	●	●	●		●		
Quick NAI		●	●	●	●		●	●	
PantaFive: Tag Weight Converter		●	●	●					
Anime-Face-Detector (Zhou, 2018/2023)						●			

*Table 5. The table provides descriptions of services that support the generation of AI images, informed by posts in the Arcalive AI Art channel, an internet forum.**

Services	**Descriptions**
Snowshell	Upscale and enhance image resolution.
Illuminarty	Show the probability that the image was generated by AI.
AI Image EXIF Viewer	Extract or suggest prompts based on image information.
NovelAI Tag Generator	Categorize tags and support completion of prompts.
WebUI Tag Autocomplete	Autocomplete tags and sort images by the number of tags in Danbooru.
AIBooru	Share AI-generated images with prompts.
Prompt Search	Share AI-generated images with prompts, sort images by each criterion, use EXIF viewer, and present the most similar images among those shared based on user-inputted prompts.
NovelAI Tag Experiments	Categorize tags and share images with prompts.
Quick NAI	Support web UI grammar, recommend tags for each user, and tag bookmarks.
PantaFive: Tag Weight Converter	Interconvert web UI and NovelAI Image Generator grammars.
Anime-Face-Detector	Automatically recognize and collect character faces from anime images.

** These services interact in two ways. First, users of the forum develop their own services for their own convenience and then promote them in posts to other users. They then interact with users in comments or posts to update and improve the service. Second, users seek out information from outside the forum and introduce it to other users in the forum through posts. There is less interaction between users inside the forum than in the first case, but some active users interact with other users outside the forum. For instance, some users visit GitHub repositories and give feedback to developers in the issues tab.*

Figure 3 is a diagram of the results of the combined analysis of the case studies and thematic analysis. When we tracked the posts made by the users in the various case studies, the majority of users who spontaneously developed the service started out as customers who looked at images and imitated the prompts. They then became explorers who searched the regularity of image generation based on the inputs and outputs, and then showed a common pattern of becoming producers who produced new information by sharing the information gained in this process with other users. Users with sufficient motivation and development skills finally became developers, and we can see from Table 1 that it is difficult for developer types to write posts, and that they invest a lot of time in individual research and development from this stage. Summarizers and innovators have sporadic postings among users who are most likely to be of each type. Unlike innovators, whose behavior is random and hard to define the cause, summarizers are driven by user motivation. As a result, we've seen talented summarizers in our forums receive administrator or moderator status and manage posts of the summarizer type for long periods of time. This is an obvious causal relationship that we've observed by analyzing the content of posts and comments, and vice versa is not. In our case study, we also observed that users tend to converge on a certain type of posting as they get later. This characteristic is related to the difficulty of the posting. For example, producer type posts often contain misinformation, and there are other users' trials in the forums to correct them. Since it is difficult to write a developer type of post, users write producer type posts, which account for the highest percentage of posts. In the long

run, as the number of users entering the forum decreases, it seems that the gap between the proportion of consumers and producers will increase, and vice versa, the gap will decrease.

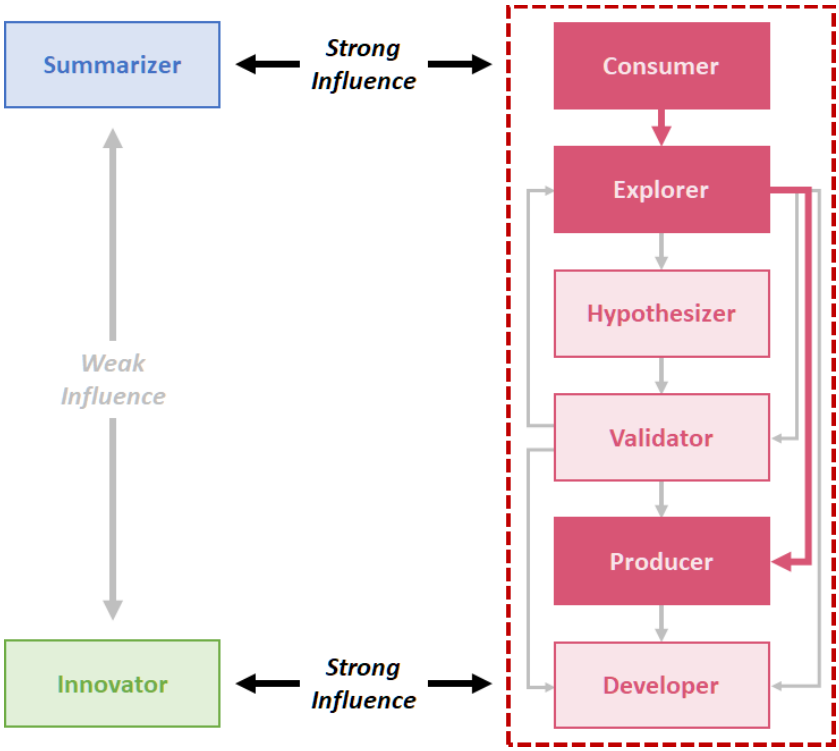

Figure 3. The figure illustrates how each type interacts and builds a relationship over time. Most users stay in the types of consumer, explorer, and producer, which are highlighted in the diagram.

4 Discussion

In this paper, Figure 3 defines the relationship between each type, and Tables 1, 2, and 4 show the behavioral tendencies of users and what types are lacking in support services. As a result, there is a need to improve the interface design in the direction of helping users evolve from producer to developer type, and there is an urgent need for support services to activate the hypothesizer type. Users also expressed the desire for a better interface for image editing with a large degree of freedom, such as retouching, in the case study. Therefore, we proposed the following for improvements.

4.1 Improve the digital drawing interface

If there are problems with the currently generated image, users rely on low-probability repetition by setting the problematic area and regenerating the image in that part. We propose digital drawing interfaces with more degrees of freedom, where colors and brushes can be used, and these drawing elements can interact with text input to edit and regenerate images. This will allow users to more easily control image generation and provide a better user experience in retouching and image regeneration.

4.2 Integrate the user interface with social networks

From Table 1, it can be inferred that users have a strong desire to perform the behavior of generating and sharing images, as the ratio of consumer and producer types is the largest. Moreover, the case studies show that users want to perform behaviors related to the developer type, but they cannot

reach the developer type due to the limitations of computing resources, difficulties in learning related knowledge, and problems in collecting training data for fine-tuning. Therefore, we propose an integrated image generation interface with social networks that can not only generate images with prompts but also share competitions and tutorials like Kaggle. We expect that having an interface design where computing resources are provided and training data is shared within one service will accelerate the fine-tuning of the models, and that user-generated tutorials will increase users reaching the developer type.

4.3 Build interaction-based recommendation systems

Most users frequently used tags such as *best quality* and *masterpiece* in prompts to improve the quality of images in the case studies, and after the introduction of the AI image EXIF viewer, we have seen the behavior of inferring and sharing tag bundles to improve the quality on specific topics. Many users have a strong desire to be recommended validated inputs. Therefore, we propose an interaction-based recommendation system, which will allow sharing of tag bundles beyond tags and a more intuitive user interface for parameter tuning. Furthermore, when using the generated images as training data, the interface can be designed in such a way that users can evaluate the quality of the images through interaction with each other on social networks, feed this information back to the AI as labels, and finally contribute to the loop to automate the construction of training data. These will allow users to control and generate the images they want using recommended tag bundles and get quantitative information about the compatibility and usability of extension programs, models, LoRA, and hypernetworks to make better choices.

4.4 Design Intuitive Style-based and Object-based Interface

From the case study, many users have shown a desire to go beyond textual representation and attempt to control and modify the style of the image and the objects within the image. Someone used to fine-tune their own models, but also to produce LoRA for post-processing by focusing on one object, such as a specific character, rather than just a style. Among posts shared by someone who had fine-tuned LoRA, it could be proven that posts shared LoRA that help generate images of the most favorite virtual characters or celebrities were getting a lot of recommendations. This means that when designing your interface, we must consider regeneration and retouching based on trials and errors, and that means we need more intuitive interfaces. Based on these needs, we propose three main interface designs. First, users can express and apply styles with single component. This can be input in the form of images, such as style images to apply style when trying style transfer, or an interface design that separately inputs the name of the desired artist or the name of the painting style. The second is an interface with a design that allows users to specify an area to edit and enter text in or around the area. This has similar experiences to the software tools that users usually use and is much more intuitive for editing. The third is an interface design where if the text and line are colored, only the components that are colored the same way interact to generate the image. This would make it easier to area within an empty image without having to use the Latent Couple Extension. It would also be nice to have a way to specify a layout with a rectangular area rather than line and could be applied to overlapping layouts by applying all the corresponding text prompts.

5 Conclusion

The use of generative artificial intelligence (AI) and its recent breakthroughs in text-to-image diffusion models are being rapidly adopted amongst creators and artists with the potential, calling for a paradigm shift. However, the users are challenged to give inputs, such as prompt engineering, to generate images that they would like to see. The current system is designed to interact with the users through text inputs. It is a challenging task for the users to create a prompt and feed it to the model to generate results that align with specific needs. To propose such human-centered design guidelines, we collected 78,911 posts from the internet community and analyzed them through the thematic analysis method. Through the analysis, we identified and classified the patterns with 25 behaviors and 8 types of users that show the collaborative process to make the best use of stable diffusion models to create the image outputs that meet the creators' requirements. We hope the guidelines can help improve the user experience, ultimately aim to build the optimal user interface based on the suggestions in the paper, which will help more creators have more effective and desired image generation processes. In terms of human-AI interaction, we also believe that it would be worthwhile as follow-up studies to introduce a tangible user interface (TUI) that provides users with easy control or to design the graphic user interface to trace and undo the branches of each image generation process to facilitate intuitive analysis. Researchers are confident that advances in AI will lead to huge changes in visual content creation. Like how the shift from analog to digital work made it easier to store and modify, we believe that the integration of AI into digital work will lead to less repetitive work and more interactions that rely on retouching based on unexpected prototypes rather than complete creation. As AI advances, it will be the mission of designers to consider human-AI interaction and refine optimal interface design to make it easier and more equitable for everyone who hopes to create.

References

Abid, A., Abdalla, A., Abid, A., Khan, D., Alfozan, A., & Zou, J. (2019). Gradio: Hassle-Free Sharing and Testing of ML Models in the Wild. arXiv. http://arxiv.org/abs/1906.02569

AIBooru. (2020). AIBooru: Anime Image Board. https://aibooru.online

Anlatan. (2022). NovelAI Image Generation. NovelAI. https://novelai.net/image

Arcalive. (2020). Arcalive AI Art Channel. https://arca.live/b/aiart

AUTOMATIC1111. (2023). Stable Diffusion web UI. GitHub. https://github.com/AUTOMATIC1111/stable-diffusion-webui

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brooks, T., Holynski, A., & Efros, A. A. (2023). InstructPix2Pix: Learning to Follow Image Editing Instructions. arXiv. http://arxiv.org/abs/2211.09800

Carlyone. (2022, October 27). NovelAI Tag Experiments. https://zele.st/NovelAI

dustysys. (2023). Detection Detailer. GitHub. https://github.com/dustysys/ddetailer

Eyal, A. (2023). Stable Diffusion Dynamic Prompts extension. GitHub. https://github.com/adieyal/sd-dynamic-prompts

hako-mikan. (2023). SuperMerger. GitHub. https://github.com/hako-mikan/sd-webui-supermerger

Illuminarty. (2022). Illuminarty: AI generated Content Detection. https://illuminarty.ai

kwaroran. (2022, November 19). Quick NAI. Chrome Web Store. https://chrome.google.com/webstore/detail/quick-nai/diikeekobcgpaggpelkjhlahhimhegoc

Lee, K. (2022, November 28). NovelAI Tag Generator. NovelAI.APP. https://novelai.app

Midjourney. (2022). Midjourney. https://www.midjourney.com/home/

OpenAI. (2022). DALL·E 2. https://openai.com/product/dall-e-2

opparco. (2023). Latent Couple Extension. GitHub. https://github.com/opparco/stable-diffusion-webui-two-shot

PantaFive. (2023, March 29). PantaFive: Tag weight Converter. https://pfive.xyz/tag/
Prompt Search. (2022). Prompt Search. https://www.ptsearch.info
Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. arXiv. http://arxiv.org/abs/2112.10752
shounksu. (2022, October 16). WebUI Tag Autocomplete. Greasy Fork. https://greasyfork.org/ko/scripts/452929-webui-%ED%83%9C%EA%B7%B8-%EC%9E%90%EB%8F%99%EC%99%84%EC%84%B1
shounksu. (2023, March 3). AI Image EXIF Viewer. Greasy Fork. https://greasyfork.org/ko/scripts/452822-ai-%EC%9D%B4%EB%AF%B8%EC%A7%80-exif-%EB%B7%B0%EC%96%B4
Tang, R., Liu, L., Pandey, A., Jiang, Z., Yang, G., Kumar, K., Stenetorp, P., Lin, J., & Ture, F. (2022). What the DAAM: Interpreting Stable Diffusion Using Cross Attention (arXiv:2210.04885). arXiv. http://arxiv.org/abs/2210.04885
Yang, Q., Scuito, A., Zimmerman, J., Forlizzi, J., & Steinfeld, A. (2018). Investigating How Experienced UX Designers Effectively Work with Machine Learning. Proceedings of the 2018 Designing Interactive Systems Conference, 585–596. https://doi.org/10.1145/3196709.3196730
Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–13. https://doi.org/10.1145/3313831.3376301
YukihoAA. (2022). waifu2x—Snowshell. GitHub. https://github.com/YukihoAA/waifu2x_snowshell
Zhang, L., & Agrawala, M. (2023). Adding Conditional Control to Text-to-Image Diffusion Models. arXiv. https://doi.org/10.48550/arXiv.2302.05543
Zhou, X. (2023). Anime-Face-Detector. GitHub. https://github.com/qhgz2013/anime-face-detector

About the Authors:

Seonuk Kim: Seonuk Kim is an undergraduate student, double majoring in Industrial Engineering and Design at UNIST. His research interests are designing intuitive web user interfaces with human-centered AI as a creative support tool.

Taeyoung Ko: Taeyoung Ko is a researcher at the Expressive Computing Laboratory, UNIST. He received his B.S. in Design from UNIST. His research interest is developing desirable and efficient web user interfaces using generative models.

Yousang Kwon: Yousang Kwon is an undergraduate student, majoring in Computer Science and Engineering at UNIST. His recent research interest is musician-computer interaction.

Kyungho Lee: Kyungho Lee is an Assistant Professor of the Design Department, leading the Expressive Computing Laboratory at UNIST. He received a Ph.D. in Informatics with an Art and Cultural Informatics concentration at the University of Illinois at Urbana-Champaign.

Acknowledgement: This work was supported by the Technology Innovation Program (20015056, Commercialization design and development of Intelligent Product-Service System for personalized full silver life cycle care) funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea).